AWARD NUMBER: W81XWH-15-1-0680

TITLE: Genetic Modeling of Radiation Injury in Prostate Cancer Patients Treated with Radiotherapy

PRINCIPAL INVESTIGATOR: Barry S. Rosenstein and Harry Ostrer

CONTRACTING ORGANIZATION: Icahn School of Medicine at Mount Sinai and Albert Einstein College of Medicine of Yeshiva University

REPORT DATE: October 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2017	Annual	30 Sep 2016 – 29 Sep 2017

4. TITLE AND SUBTITLE

Genetic Modeling of Radiation Injury in Prostate Cancer Patients Treated with Radiotherapy

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-15-1-0680

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Barry S. Rosenstein and Harry Ostrer

E-Mail: barry.rosenstein@mssm.edu; harry.ostrer@einstein.yu.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Icahn School of Medicine at Mount Sinai
Dept. of Radiation Oncology
Box 1236,One Gustave Levy Place
New York, NY 10029

Albert Einstein College of Medicine
1300 Morris Park Avenue
Ullmann Building, Room 817
Bronx, NY 10461

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

From the completed analyses performed during the first year of the project, we have identified eleven SNPs that show an association with two-year toxicity following prostate cancer radiotherapy. Three of these single nucleotide polymorphisms (SNPs) meet the stringent threshold for genome-wide significance (meta-p-value < 5x10-8), and eight others approached genome-wide significance. For the three genome-wide significant SNPs, we found that the direction of the effect was consistent across all studies for which data were available.

15. SUBJECT TERMS

Radiogenomics, single nucleotide polymorphisms, prostate cancer, radiation therapy, adverse effects, urinary morbidity, rectal injury, sexual dysfunction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	16	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. INTRODUCTION

As with all forms of treatment for prostate cancer, the goal of radiotherapy is to provide patients with a sustainable cure of their tumor without causing substantial damage to normal tissues and organ function. Clearly, there have been great advances to conform the radiation field to the cancer. However, even with dosimetric improvements, some volume of normal tissue still receives a substantial radiation dose during the course of radiotherapy. This radiation exposure often results in toxicity that compromises organ function and affects the quality of life for the prostate cancer survivor. Therefore, an important goal is to create an assay that could predict which patients are most likely to develop radiation-induced complications. The main approach taken in recent years to achieve this goal has been the identification of genetic markers, primarily single nucleotide polymorphisms (SNPs), which are associated with the development of adverse effects resulting from radiotherapy. The aim of this research is to identify the genetic markers that can serve as the basis for personalized radiotherapy in which cancer management is formulated so that it optimizes the treatment plan for each patient based upon their genetic background. The overall objective of this research project is to create a robust, validated, sensitive and specific SNP-based assay that will be ready for implementation in the clinical setting. This assay will be capable of predicting the risk of developing adverse effects resulting from radiotherapy treatment of prostate cancer -- erectile dysfunction, urinary morbidity and rectal injury. The purpose of the current project is to validate previously identified SNPs and to discover new SNPs in a large, independent cohort and to develop a predictive instrument and companion diagnostic.

2. KEYWORDS:

Radiogenomics, single nucleotide polymorphisms, prostate cancer, radiation therapy, adverse effects, urinary morbidity, rectal injury, sexual dysfunction

3. ACCOMPLISHMENTS:

What were the major goals of the project?

- *Validate previously discovered SNPs and identify additional SNPs via meta-analysis of GWAS using a substantially expanded set of studies in which approximately 7,000 men treated with radiotherapy for prostate cancer have been genotyped using a SNP array that contains a set of genome-wide SNPs as well custom content that contains our previously identified SNPs. (Months 1-18).*

This represented the major goal for the first year of the project. The results were outlined in the annual report submitted last year with additional details for additional work that was accomplished during the second year of the project provided below.

- *Create polygenic risk models from results of single-SNP analysis and investigate effects of demographic, dosimetric and clinical factors on polygenic risk models. (Months 12-30).*

This represents the major goal for the second year of the project, the results of which are described below.

- *Use cross-validation to obtain accurate effect sizes and estimates of sensitivity and specificity (Months 25-30)*

This represents an important goal for the third year of the project.

- *Develop a low-cost, high-performance genetic assay (Months 1-34)*

Efforts to achieve this goal were initiated as outlined below.

- *Export the models developed in Aim 2 to a web-based application that could be used by physicians in practice and/or genetic testing laboratories. (Months 24-36)*

This represents a major goal for the final six months of the project.

What was accomplished under these goals?

KEY RESEARCH ACCOMPLISHMENTS:

Completion of GWAS meta-analysis

During the second year of the funding period, we completed the meta-analysis of genome-wide association studies (GWAS) of late radiotherapy toxicity from the Radiogenomics Consortium. We used two different analytic approaches in the GWAS meta-analysis: 1) logistic regression to test association of each SNP with grade 1 or worse toxicity at 2 years post-radiotherapy, and 2) survival analysis to test association of each SNP with cumulative incidence of grade 2 or worse toxicity considering all assessments between 6 months and 5 years post-radiotherapy.

The first approach, analysis of 2-year toxicity prevalence, is the primary approach described in the grant proposal and is the approach used in our pilot study [PMID 27515689] in which we performed a meta-analysis of a subset of the GWAS datasets that were available prior to the Oncoarray genotyping initiative. The rationale for this approach is that the vast majority of participants in the various GWAS datasets had a minimum of 2 years of follow-up for toxicity following radiotherapy, and thus prevalence of toxicity at this time point was a simple and unbiased endpoint. The prevalence of toxicity in each GWAS cohort is shown in Table 1, along with covariates that were included in the logistic regression model testing association between every SNP and each toxicity endpoint. In addition to these four tissue and symptom-specific endpoints, we also analyzed overall toxicity measured by STAT score [PMID 21605943]. Table 2 lists the endpoints included in analysis of STAT score in each study, and Figure 1 shows the distribution of STAT scores in each cohort.

Every SNP was tested for association with each 2-year toxicity prevalence endpoint using either logistic regression (binary endpoints) or linear regression (STAT score), adjusting for covariates of importance that captured patient or treatment heterogeneity across studies. Quantile-quantile plots (Figure 2) exhibited no evidence of genomic inflation, suggesting that ancestry was well controlled by limiting analysis to individuals of European ancestry as determined by principle components analysis. We selected SNPs with meta-p-values < 0.1 to carry forward in polygenic risk score modeling described below. Unfortunately no single SNP reached the stringent threshold for genome-wide significance after filtering out SNPs with minor allele frequency less than 5%. This filter was applied based on our *a priori* statistical power calculation. SNPs that are more rare can result in poor model fit with our given sample size, leading to spurious associations.

Previous GWAS and candidate gene studies of prostate radiotherapy toxicity have identified several SNPs showing an association with toxicity at 2 years or STAT score, and we evaluated these SNPs for validation in the present GWAS meta-analysis. As reported in Table 3, we were able to validate the four prior SNP-toxicity associations.

The secondary analytic approach, time-to-event analysis of each toxicity outcome, addresses limitations of the 2-year prevalence analysis. The main limitation of analyzing 2-year prevalence is that this approach ignores toxicity data collected prior to and after the 2-year time-point. Previous studies suggest that late toxicity following radiotherapy for prostate cancer can develop many years after radiotherapy, and thus we are likely missing toxicity ‘cases’ by considering only 2-year assessment. Therefore, we performed a secondary GWAS meta-analysis using time to onset of toxicity, considering all follow-up assessments from 6 months to 5 years post-radiotherapy, censoring individuals who did not reach 5 years of follow-up at their last recorded visit. Because this approach uses a larger proportion of the data, we were able to dichotomize the toxicity measures using a more clinically meaningful cut-point (grade 2 or worse toxicity) compared with the analysis of 2-year prevalence. The cumulative incidence of grade 2 or worse toxicity for each study is provided in Figure 3. As expected, toxicity continues to develop after 2 years, highlighting the importance of considering all follow-up assessments. As expected, urinary toxicity was more incident in the GenePARE cohort, for which patients were treated with brachytherapy with or without additional external beam radiotherapy.

The Cox proportional hazards model was used to test for association between SNPs and time-to-onset of grade 2 or worse toxicity. We used interval censoring and the Efron method of breaking ties. Using this

approach, we were successful in identifying several loci that reached genome-wide significance. Figure 4 shows Manhattan plots for each toxicity endpoint, and Table 4 lists SNPs tagging the significant loci. A total of six loci reached significance: chr5q33.3 (OR 1.99; 95% CI 1.61-2.47; meta-p-value $3.14x10^{-10}$) and chr8q24.23 (OR 2.17; 95% CI 1.64-2.86; meta-p-value $2.12x10^{-8}$) associated with rectal bleeding, chr3q29 (OR 1.92; 95% CI 1.54-2.44; meta-p-value $3.22x10^{-8}$) and 9p21.1 (OR 3.85; 95% CI 2.52-5.89; meta-p-value $4.71x10^{-10}$) associated with decreased urine stream, and 1q42.2 (OR 1.93; 95% CI 1.52-2.43; meta-p-value $2.29x10^{-8}$) and 3p14.3 (OR 4.99; 95% CI 2.83-8.80; meta-p-value $2.69x10^{-8}$) associated with hematuria. A seventh locus on chromosome 10 was associated with rectal bleeding, but the odds ratios and standard errors in the individual studies were large, indicating a poorly fit model and a likely spurious association.

Table 1. Prevalence of grade 1 or worse toxicity at 2 years following radiotherapy (RT) among each GWAS cohort. Toxicity cases are men with any grade 1 or worse toxicity at 2 years; toxicity controls are men with grade 0 toxicity at 2 years.

Study (N)	Toxicity cases, N(%)	Toxicity controls, N(%)	Covariates included in logistic regression models
INCREASED URINARY FREQUENCY[a]			
RAPPER[b] (N = 1,876)	289 (15.4%)	1,587 (84.6%)	pre-RT daytime frequency, pre-RT nocturia, age, total BED, genotyping batch
RADIOGEN[c] (N = 597)	89 (14.9%)	508 (85.1%)	age, total BED, hormones, surgery, TURP
GenePARE[d] (N = 398)	122 (30.7%)	276 (69.4%)	pre-RT daytime frequency, pre-RT nocturia, age, total BED, hormones, genotyping batch
Ghent (N = 281)	33 (11.7%)	248 (88.3%)	pre-RT daytime frequency, pre-RT nocturia, age, total BED, hormones, surgery, TURP
CCI-EBRT[e] (N = 148)	22 (14.9%)	126 (85.1%)	age, total BED, hormones, TURP
DECREASED URINE STREAM[a,f]			
RAPPER[b] (N = 1,937)	112 (5.8%)	1,825 (94.2%)	pre-RT retention, age, total BED, genotyping batch
RADIOGEN[c] (N = 602)	5 (0.8%)	597 (99.2%)	age, total BED, hormones, surgery
GenePARE[d] (N = 345)	102 (29.6%)	243 (70.6%)	pre-RT weak stream, age, total BED, hormones, genotyping batch
HEMATURIA[a,g]			
RAPPER[b] (N = 1,990)	26 (1.3%)	1,964 (98.7%)	age, total BED, genotyping batch
RADIOGEN[c] (N = 597)	16 (2.7%)	581 (97.3%)	age, total BED, hormones, surgery, TURP
GenePARE[d] (N = 495)	17 (3.4%)	478 (96.6%)	age, total BED, hormones, genotyping batch
Ghent (N = 280)	9 (3.2%)	271 (96.8%)	age, total BED, hormones, surgery
RECTAL BLEEDING[h]			
RAPPER[b] (N = 1,946)	260 (13.4%)	1,686 (86.6%)	age, total BED, diabetes, genotyping batch
RADIOGEN[c] (N = 600)	71 (11.8%)	529 (88.2%)	age, total BED, diabetes, hormones, surgery
Ghent (N = 277)	22 (7.9%)	255 (92.1%)	age, total BED, diabetes, hormones, surgery
CCI-BT	11 (7.5%)	136 (92.5%)	age, total BED, diabetes, hormones

(N = 147)			
CCI-EBRT[e] (N = 145)	26 (17.9%)	119 (82.1%)	age, total BED, diabetes, hormones

[a] Urinary endpoints were not available for analysis if the CCI-BT cohort
[b] All RAPPER participants received hormone therapy and none received prior surgery; prior TURP was not available
[c] Toxicity grading in RADIOGEN accounts for baseline symptoms, and so the baseline score is not needed in the model
[d] None of the participants in GenePARE received prior surgery
[e] None of the participants in CCI-EBRT received prior surgery. Toxicity grading in CCI-EBRT accounts for baseline symptoms, and so the baseline score is not needed in the model
[f] Decreased stream at 2 years was too rare to be analyzed in the CCI-EBRT and UGhent cohorts
[g] Hematuria at 2 years was too rare to be analyzed in the CCI-EBRT
[h] Rectal bleeding at 2 years was not available in GenePARE

Table 2. Toxicity endpoints included in calculation of STAT score, by study[a]. For each endpoint, the worst score from between 2 and 5 years post-radiotherapy was used to calculate STAT.

Endpoint	**RAPPER (N = 1,979)**	**RADIOGEN (N = 603)**	**GenePARE (N = 462)**	**UGhent (N = 281)**	**CCI-EBRT (N = 145)**
Rectal bleeding	✔	✔	✔	✔	✔
Diarrhea	✔	✔		✔	✔
GI incontinence	✔	✔			✔
Proctitis	✔	✔		✔[b]	✔
Urinary frequency	✔	✔	✔	✔	✔
Cystitis	✔	✔	✔[c]	✔	✔
Urinary retention	✔	✔	✔		✔
Urinary incontinence	✔	✔	✔[d]	✔	✔

[a] STAT was not calculated in CCI-BT because only a single toxicity endpoint (rectal bleeding) was assessed.
[b] Rectitis was assessed in the UGhent cohort
[c] Hematuria was assessed in the GenePARE cohort
[d] Urinary urgency was assessed in the GenePARE cohort

Table 3. Validation of previously reported toxicity risk loci.

Locus[a]	Original Publication	MAF[b]	RGC GWAS meta-analysis		
			Endpoint	OR (95% CI)[c]	p-value
rs1801516 (*ATM*) Chr11:108,175,462	$STAT_{acute}$ and $STAT_{late}$ in prostate and breast patients [PMID 27443449]	0.22	$STAT_{late}$	0.043 (0.010, 0.076)	0.011
rs264663 (*TANC1*) Chr2:159,910,206	$STAT_{late}$ in prostate patients [PMID 24974847]	0.05	$STAT_{late}$	0.150 (0.046, 0.254)	$4.88x10^{-3}$
rs7720298 (*DNAH5*) Chr5:13,858,328	2yr Decreased Stream in pilot GWAS meta-analysis [PMID 27515689]	0.30	2yr Decreased Stream	1.36 (1.08, 1.71)	$8.44x10^{-3}$
rs17599026 (*KDM3B*) Chr5:137,763,798	2yr Urinary Frequency in pilot GWAS meta-analysis [PMID 27515689]	0.07	2yr Urinary Frequency	1.51 (1.21, 1.89)	$3.40x10^{-4}$

[a] Base position from GRCh37/hg19
[b] Minor allele frequency, from PRACTICAL Oncoarray samples of European ancestry
[c] Beta coefficient in the case of STAT score

Figure 1. Distribution of STAT score in each GWAS cohort.

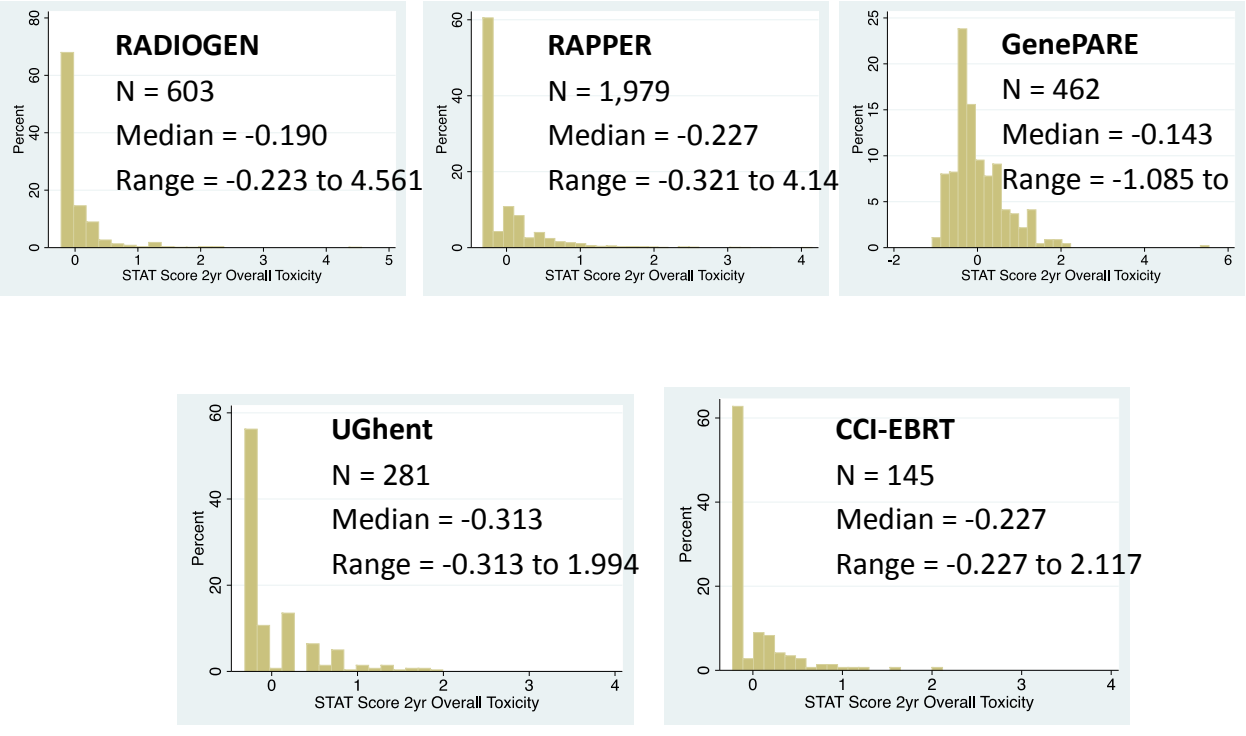

Figure 2. Quantile-quantile plots of each 2-year toxicity prevalence endpoint.

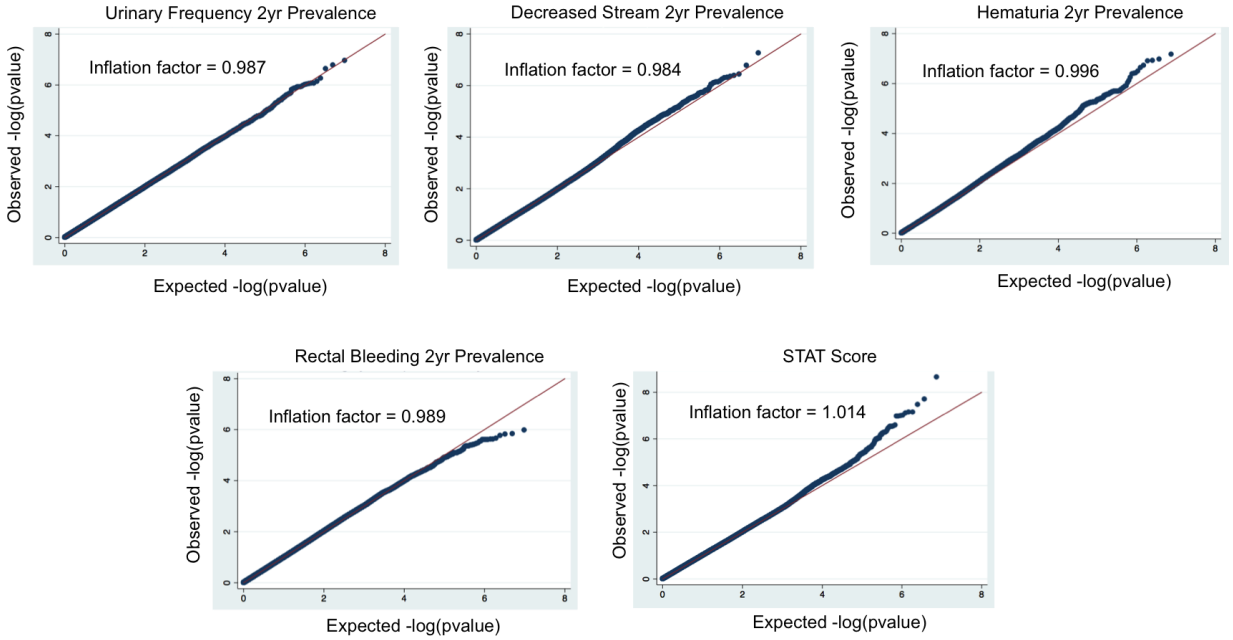

Figure 3. Cumulative incidence of grade 2+ toxicity following radiotherapy.

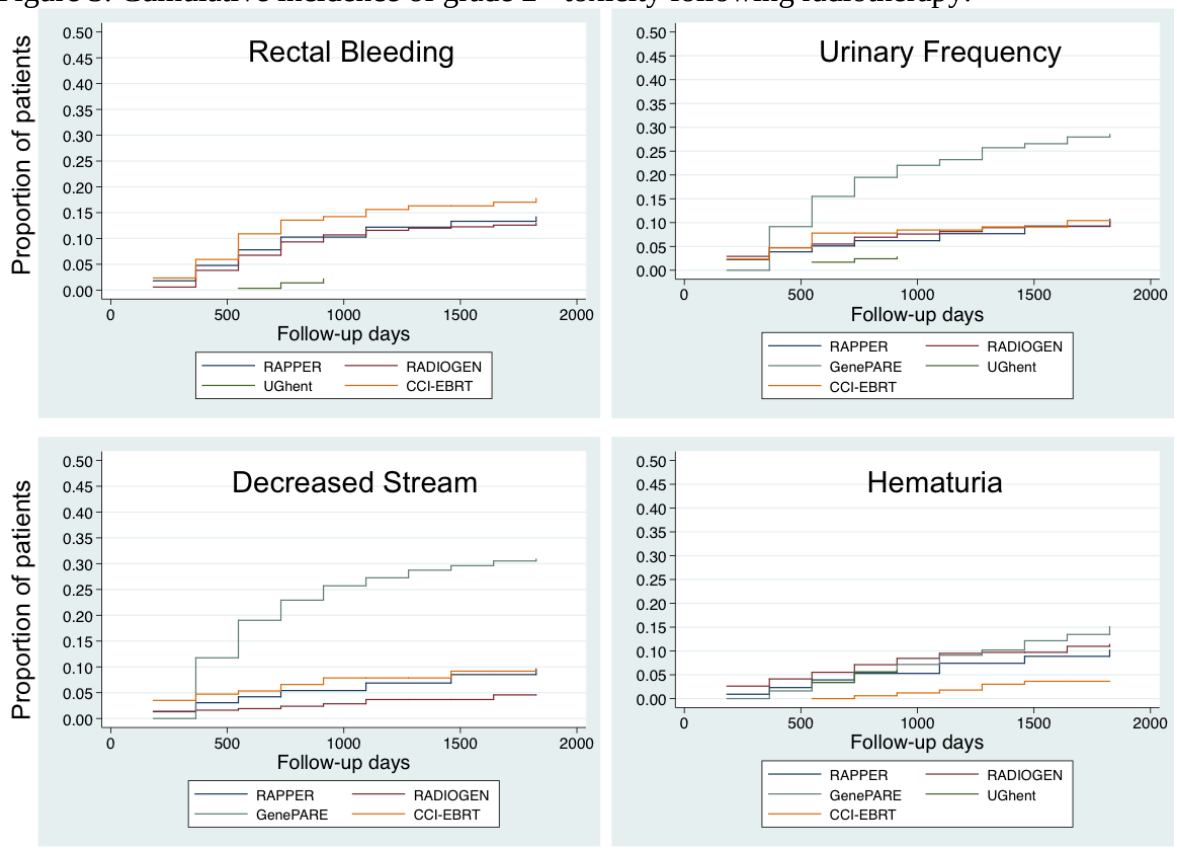

Figure 4. Manhattan plots showing results of GWAS meta-analysis of Cox regression of time-to-onset of grade 2 or worse toxicity.

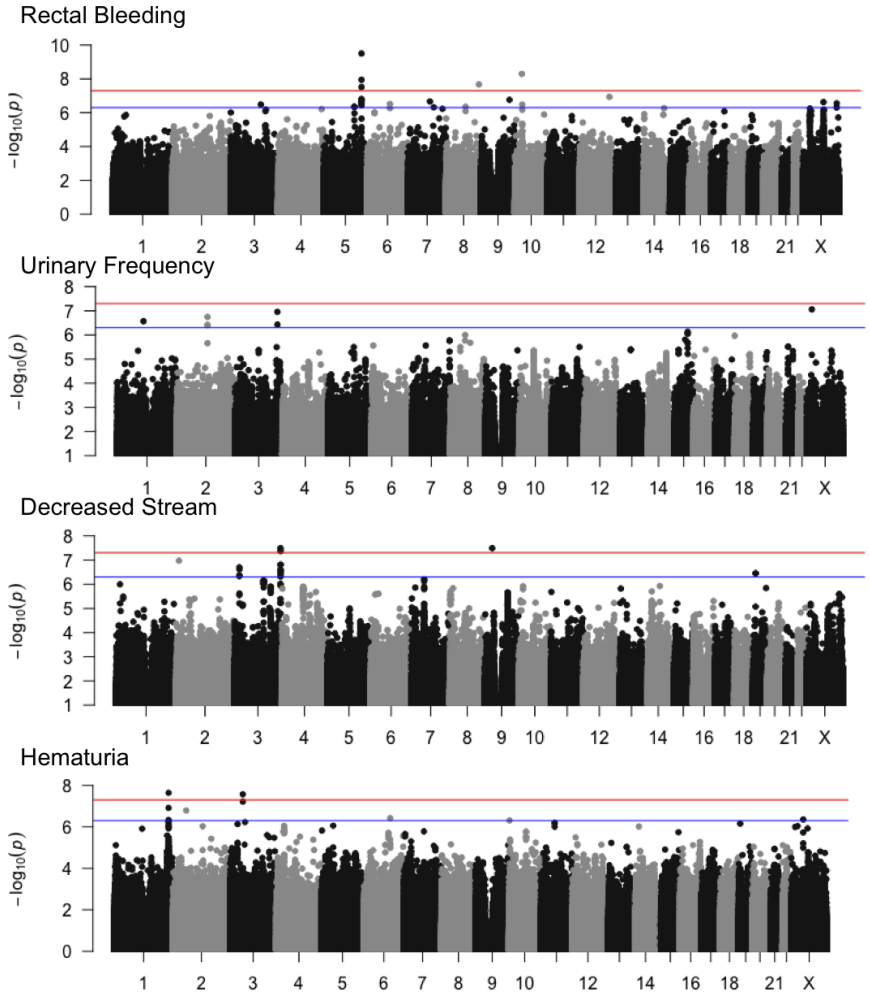

Table 4. Genome-wide significant loci associated with incidence of grade 2 or worse radiotherapy toxicity.

Locus	Toxicity	Study	R-sq	HR (95% CI)	p-value
Chr5:157403410:A:G MAF = 0.093	Rectal bleeding	RAPPER	0.986	1.84 (1.40, 2.42)	
		RADIOGEN	0.986	2.58 (1.69, 3.95)	
		UGhent	0.986	1.38 (0.18, 10.4)	
		CCI-EBRT	0.979	1.27 (0.38, 4.25)	
		CCI-BT	0.986	2.01 (0.97, 4.20)	
		Meta-analysis		1.99 (1.61, 2.47)	3.14×10^{-10}
Chr8:137163144:C:T MAF = 0.046	Rectal bleeding	RAPPER	0.777	1.47 (1.00, 2.17)	
		RADIOGEN	0.777	4.17 (2.70, 6.67)	
		UGhent	0.777	1.54 (0.11, 25.0)	
		CCI-EBRT	0.833	1.20 (0.36, 4.00)	
		CCI-BT	0.777	1.39 (0.52, 3.70)	
		Meta-analysis		2.17 (1.64, 2.86)	2.12×10^{-8}
Chr9:30868163:T:C MAF = 0.048	Decreased Stream	RAPPER	0.947	1.73 (0.71, 4.20)	
		RADIOGEN	0.947	2.03 (0.27, 15.40)	
		GenePARE	0.947	4.36 (2.55, 7.46)	
		CCI-EBRT	0.945	14.34 (3.78, 54.4)	
		Meta-analysis		3.85 (2.52, 5.89)	4.71×10^{-10}
Chr:1:230837180:C:T MAF = 0.056	Hematuria	RAPPER	0.995	1.40 (0.96, 2.04)	
		RADIOGEN	0.995	2.40 (1.54, 3.73)	

		UGhent	0.995	3.59 (1.72, 7.49)	
		GenePARE	0.995	2.01 (1.25, 3.22)	
		CCI-EBRT	1.00	0.99 (0.13, 7.58)	
		Meta-analysis		1.93 (1.53, 2.43)	$2.29x10^{-8}$
Chr:3:54729912:C:T MAF = 0.042	Hematuria	RAPPER	0.737	3.46 (1.66, 7.21)	
		RADIOGEN	0.737	12.46 (4.50, 34.5)	
		UGhent	0.737	5.69 (0.61, 52.9)	
		GenePARE	0.737	0.42 (0.02, 11.5)	
		CCI-EBRT	NA	NA	
		Meta-analysis		4.99 (2.83, 8.80)	$2.69x10^{-8}$

Polygenic Scores methods

Polygenic Score (PGS) is a quantitative summary of genetic predisposition of a certain phenotypic traits. In this study, we constructed PGS on five traits: type 2 diabetes (T2D), Decstrm2yr, Hematuria2yr, Recbld2yr, UrineFreq2yr. T2D served as a positive control trait since large GWAS for T2D have been reported and summary statistic data are available. The four radiation toxicity endpoints are the main focus of the analysis. GWAS summary data (termed "training data") were used in constructing the PGS formula, which is a linear combination of selected variants. For the T2D PGS formula, we used DIAGRAM study results [PMID 22885922]. For the radiation toxicity PGS formula, we used the GWAS meta-analysis results based on all GWAS cohorts described above (RAPPER, RADIOGEN, GenePARE, UGhent, CCI-EBRT, and CCI-BT). Then the PGS formula was applied to MSSM and RAPPER cohort to compute the PGS score for the five traits, denote as PGS_{T2D}, $PGS_{Decstrm2yr}$, $PGS_{Hematuria2y}$, $PGS_{Recbld2yr}$ and $PGS_{UrineFreq2yr}$. In brief detail, 9,621,254 variants available on MSSM and RAPPER cohorts were used in the analysis. For each of the 5 GWA endpoints, we then proceeded as follows: (1) align "training data" alleles to the 1000 Genome Reference (hg19), and adjust beta coefficients accordingly; (2) subset training data SNPs to the list of variants shared by MSSM and RAPPER cohorts; (3) prune variants based on the 1000G EUR cohort linkage dis-equilibrium (LD), to remove tightly correlated variants; (4) filter the pruned variant list by GWA p-value threshold (e.g. 10^{-3}); (5) compute the five PGS (PGS_{T2D}, $PGS_{Decstrm2yr}$, $PGS_{Hematuria2}$y, $PGS_{Recbld2yr}$ and $PGS_{UrineFreq2yr}$) on each MSSM and RAPPER subjects; lastly, we tested the association between PGS and observed values of the five traits using logistic regression models.

Polygenic Scores Results

In table 5 we report the estimated associations between polygenic score of five traits computed using a p-value threshold of 1E-3, and corresponding observed traits. Firstly, x1 the PGS_{T2D} based on DIAGRAM study significantly associated with diabetes status in both RAPPER and MSSM, validating the analytical pipeline. Further, all $PGS_{Decstrm2yr}$, $PGS_{Hematuria2}$y, $PGS_{Recbld2yr}$ and $PGS_{UrineFreq2yr}$ significantly associated with observed traits. For example, $PGS_{Decstrm2yr}$ associated with observed decstrm_2yr in the RAPPER and MSSM cohorts with p-values of 4.28E-14 and 2.87E-13, respectively. The significant associations in Table 5 suggest that genetic factors (summarized as polygenic score) have great prediction value for radiation toxicity endpoints.

Table 5. Association between polygenic score and observed trait value

endpoint	# SNPs	cohort	log-OR	std.error	T-statistic	p.value	sample size
Diabetes	1500	RAPPER	0.42	0.071	5.93	3.00×10^{-09}	2217
		MSSM	0.35	0.190	1.84	6.54×10^{-02}	617
decstrm_2yr	1956	RAPPER	2.85	0.377	7.55	4.28×10^{-14}	2068
		MSSM	6.43	0.881	7.30	2.87×10^{-13}	358
hematuria_2yr	4317	RAPPER	0.88	0.246	3.57	3.58×10^{-04}	2121
		MSSM*	*+Inf	*NA	*NA	*NA	*515
recbld_2yr	1965	RAPPER	3.42	0.224	15.26	1.32×10^{-52}	2097
		MSSM	-	-	-	-	-
urinefreq_2yr	1843	MSSM	2.81	0.293	9.61	7.36×10^{-22}	411
		RAPPER	3.13	0.166	18.83	4.33×10^{-79}	2126

* In the MSSM cohort, for the 'hematuria_2yr' endpoint we observed perfect separation between successes and failures according to the PGS, thus the logistic regression could not converge.

Machine learning methods and results

Given the promising results from the meta-analyses and polygenic score modeling, our next effort was to use machine learning (ML) to develop a panel of SNPs that can collectively classify/predict radiotherapy toxicity endpoints as accurately as possible. In previous work (http://www.biorxiv.org/content/early/2017/07/10/145771), some members of our team developed an ML pipeline that combined methods from feature selection (PMC 17720704), classification (http://www.cs.waikato.ac.nz/ml/weka/book.html) and statistical analysis (http://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf) to develop a similar gene expression-based panel for asthma diagnosis. The pipeline also includes procedures to control for the potential adverse effects of common challenges faced with ML analyses, such as model overfitting and imbalance of class sizes. The panel found using this pipeline doesn't only accurately classify asthma status, but it is based on the expression of just ninety genes, thus making the clinical translation and deployment of the pipeline more economically and practically feasible.

We have begun working on adapting the above pipeline to develop a similar SNP-based panel for predicting radiotherapy toxicity endpoints. This adaptation will incorporate the statistics developed in the above meta-analyses and polygenic score modeling to make the pipeline more relevant for SNP data. We will initially follow the same design for development of the panel(s) and their subsequent validation to be accomplished using separate cohorts to ensure fairness of the results obtained. Also, we will initially focus on the Recbld2yr and UrineFreq2yr endpoints as they have a manageable imbalance between cases and controls. However, as the work progresses, we will also investigate other designs and endpoints.

Develop a low-cost, high-performance genetic assay.

Previously, assays were developed using the quantitative polymerase chain reaction (qPCR), digital polymerase chain reaction (dPCR), and NextGen Genotyping platforms using candidate variants. Given the number of variants that we will choose to identify to test polygenic risk score and machine learning models, we have tested hybrid capture NextGen sequencing methods and have found very reproducible results with a standard set of samples.

What opportunities for training and professional development has the project provided?

Nothing to Report

How were the results disseminated to communities of interest?

Results of these findings were presented at the annual Radiogenomics Consortium Meeting in Barcelona, Spain on June 19, 2017.

What do you plan to do during the next reporting period to accomplish the goals?

An important task for the next reporting period will be to continue development of polygenic risk models from results of single-SNP analysis. We will then employ a cross-validation strategy, as well as independent test cohorts, to evaluate the prediction models created in terms of their individual predictive accuracy, sensitivity, specificity as well as the overall ROC curve. Cross-validation will indicate the most effective approach(es) to predict toxicity levels from the available SNP and clinical data. A major goal of the final year of the project will be to export the models developed in this study to a web-based application that could be used by physicians in practice and/or genetic testing laboratories. As part of this aim, we will create and test the web-based tool to assess accuracy and correct any bugs prior to making it publically available.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?
Nothing to Report

What was the impact on other disciplines?
Nothing to Report

What was the impact on technology transfer?
Nothing to Report

What was the impact on society beyond science and technology?
Nothing to Report

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change
Nothing to Report

Actual or anticipated problems or delays and actions or plans to resolve them
Nothing to Report

Changes that had a significant impact on expenditures
Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects
Nothing to Report
Significant changes in use or care of vertebrate animals.
Nothing to Report
Significant changes in use of biohazards and/or select agents
Nothing to Report

6. PRODUCTS:
Nothing to Report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: Harry Ostrer
Project Role: co-PI
Researcher Identifier: 0000-0002-2209-5376
Nearest person month worked: 1
Contribution to Project: Dr. Ostrer oversaw the design and management of this study and worked to develop assays that could be used for risk assessment.
Funding Support: This award

Name: Kinnari Upadhyay
Project Role: Bioinformatician
Researcher Identifier : N/A
Nearest person month worked: 6
Contribution to Project: Ms. Upadhyay developed a database and risk assessment tools for incorporation of genetic data for this project under the supervision of Dr. Ostrer.
Funding Support: This award

Name: Johnny Loke
Project Role: Research associate
Researcher Identifier : N/A
Nearest person month worked: 2
Contribution to Project: Mr. Loke developed qPCR, dPCR, AmpliSeq and hybrid capture sequencing assays for analysis of genetic variants identified in this project under the supervision of Dr. Ostrer.
Funding Support: This award

Name: Ke Hao
Project Role: Co-Investigator
Researcher Identifier : NA
Nearest person month worked: 2
Contribution to Project: Design and implement algorithms in constructing and evaluating polygenic score (PGS) on radiation toxicity traits.
Funding Support: This award

Name: Antonio Di Narzo, PhD
Project Role: Data analyst
Researcher Identifier : NA
Nearest person month worked: 2 months
Contribution to Project: polygenic score data analysis
Funding Support: NA

Name: Gaurav Pandey
Project Role: Co-Investigator
Researcher Identifier : NA
Nearest person month worked: 1
Contribution to Project: Design of machine learning strategies to identify genetic predictors of radiotoxicity
Funding Support: This award

Name: Mehmet Eren Ahsen

Project Role: Data Analyst
Researcher Identifier : NA
Nearest person month worked: 1
Contribution to Project: Implementation of machine learning strategies to identify genetic predictors of radiotoxicity
Funding Support: This award

Name: Barry Rosenstein
Project Role: Principal Investigator
Researcher Identifier : NA
Nearest person month worked: 2
Contribution to Project: Worked with Dr. Kerns to obtain and harmonize dosimetric, clinical and OncoArray genotyping data for all subjects from each cohort comprising this project and to perform statistical analysis for validation of previously discovered SNPs and identification of new SNPs. Worked with Drs. Pandey and Hao to use novel strategies for radiogenomics, sparse learning, polygenic score and ensemble learning, to create polygenic risk models to predict the incidence of radiotherapy toxicity based on the genotype and clinical characteristics.
Funding Support: This award

Name: Hindy Korenblit
Project Role: Data Manager
Researcher Identifier: NA
Nearest person month worked: 4
Contribution to Project: Worked with Dr. Rosenstein to organize the anonymized clinical data for the Mount Sinai cohort included in this study.
Funding Support: This grant

Name: Sarah Kerns
Project Role: Co-investigator
Researcher Identifier : NA
Nearest person month worked: 5
Contribution to Project: Dr. Kerns performed data management and statistical analyses for the GWAS meta-analysis to identify SNPs associated with radiation toxicity in collaboration with Drs. Rosenstein and Ostrer.
Funding Support: NCI K07 CA187546

Name: Andrea Baran
Project Role: Biostatistician
Researcher Identifier (e.g.ORCID ID): NA
Nearest person month worked: 1
Contribution to Project: Ms. Baran assisted with performing quality checks and data cleaning for the oncoarray SNP datasets analyzed in this project under the supervision of Dr. Kerns.
Funding Support: NCI K07 CA187546 and SBIR HHSN261201500043C

Name: Ashley Amidon Morlang
Project Role: Study Coordinator
Researcher Identifier (e.g.ORCID ID): NA
Nearest person month worked: 1
Contribution to Project: Ms. Morlang assisted with data management related to the clinical and dosimetric data for each cohort included in the GWAS analysis under the supervision of Dr. Kerns. She also coordinated the IRB exemption request/approval required for this project.
Funding Support: This award

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
Nothing to Report

What other organizations were involved as partners?
Nothing to Report